Incoherent Frequency Fusion for Broadband Steered Response Power Algorithms in Noisy Environments

Daniele Salvati, Carlo Drioli, *Member, IEEE*, and Gian Luca Foresti, *Senior Member, IEEE*

***Abstract*—The steered response power (SRP) algorithms have been shown to be among the most effective and robust ones in noisy environments for direction of arrival (DOA) estimation. In broadband signal applications, the SRP methods typically perform their computations in the frequency-domain by applying a fast Fourier transform (FFT) on a signal portion, calculating the response power on each frequency bin, and subsequently fusing these estimates to obtain the final result. We introduce a frequency response incoherent fusion method based on a normalized arithmetic mean (NAM). Experiments are presented that rely on the SRP algorithms for the localization of motor vehicles in a noisy outdoor environment, focusing our discussion on performance differences with respect to different signal-to-noise ratios (SNR), and on spatial resolution issues for closely spaced sources. We demonstrate that the proposed fusion method provides higher resolution for the delay-and-sum SRP, and improved performances for minimum variance distortionless response (MVDR) and multiple signal classification (MUSIC).**

***Index Terms*—Broadband steered response power, incoherent frequency fusion, normalized arithmetic mean, direction of arrival estimation, microphone array.**

I. Introduction

THE steered response power (SRP) algorithms are widely used for estimating the direction of arrivals (DOAs) in far-field conditions, which is a crucial step in a localization system. An important DOA application addressed in this paper involves the multiple acoustic sources localization in outdoor noisy environments for audio surveillance and scene analysis. The SRP is based on maximizing the power output of a beamformer. SRP algorithms have been developed for narrowband signals, and several methods have been proposed for wideband signals. Typically, broadband SRP is computed in the frequency-domain by applying a fast Fourier transform (FFT) on a portion of the signal and by calculating the response power on each frequency bin. Subsequently, a fusion of these estimates is computed and the estimation of the DOAs of acoustic sources is obtained by searching the local maxima on the response power spectrum. The fusion of narrowband SRP can be obtained by incoherent or coherent averaging with respect to frequency.

The delay-and-sum SRP [1] is typically computed on wideband signals by calculating an incoherent arithmetic mean (AM) average of the contributions of the microphone array. Unfortunately, the spatial resolution of SRP is poor, because the response power function is characterized by large peaks, and this makes its application unsuitable for a multi-source scenario. An advantage of using the SRP with the phase transform (PHAT) weighting function [2] is that it provides narrower response power peaks (since it reduces the auto-correlation effect), thus increasing the spatial resolution and permitting the estimation of DOAs for multiple sources [3]. The minimum variance distortionless response (MVDR) filter is based on the narrowband adaptive Capon beamformer [4]. In [5], three wideband MVDR algorithms are discussed, and the authors demonstrate the better performance of MVDR with the incoherent geometric mean (GM) if compared with AM and harmonic mean. Finally, the multiple signal classification (MUSIC) algorithm is another high resolution beamforming technique developed for narrowband signals [6], and based on an eigensubspace decomposition method. Broadband MUSIC has been proposed with incoherent signal subspace processing [7], and with coherent wideband methods [8]–[10]. In [8], [9], algorithms are proposed that require to find a focusing matrix, which allows for a proper alignment of spatial data covariance matrix. However, the estimation performance of these algorithms heavily depends on the initial conditions selected for the focusing matrix computation. In [10], the proposed method does not require any initial values to find focusing matrices, but it has an optimal performance only for moderate signal-to-noise ratio (SNR) values.

Incoherent averaging effectiveness decreases when the SNR at each frequency bin varies, since the DOA estimate at some frequencies may be affected by large errors, and the final frequency data combination may be inaccurate. Besides that, the GM based algorithms, which perform best for narrowband responses with wide numeric ranges, suffer from performance drop when the narrowband SRP presents near to zero values with consequent reduction of power peaks intensity. To mitigate these problems, we introduce an incoherent frequency combination based on a normalized arithmetic mean (NAM), which has the advantage of enhancing robustness by weighting the function values calculated on each frequency bin, so that each response power contributes equally to the final value of fusion. We demonstrate that the proposed method improves the performance of SRP, MVDR and MUSIC algorithms.

II. Broadband Steered Response Power

We assume N acoustic sources and an array composed of M microphones, and assume omnidirectional characteristics for both the sources and the microphones. The discrete-time signal received by the mth microphone can be modeled, for a free-field environment, as

$$x_m(k) = \sum_{n=1}^{N} \alpha_{nm} s_n(k - k_n - \tau_{nm}) + v_m(k) \qquad (1)$$

D. Salvati, C. Drioli, and G.L. Foresti are with the Department of Mathematics and Computer Science, University of Udine, Udine 33100, Italy, e-mail: daniele.salvati@uniud.it, carlo.drioli@uniud.it, gianluca.foresti@uniud.it.

where α_{nm} is the attenuation of the sound propagation (inversely proportional to the distance from source n to microphone m), $s_n(k)$ are the unknown uncorrelated source signals, k_n is the propagation time from the unknown source n to the reference sensor of the array, τ_{nm} is the time difference of arrival (TDOA) between the mth microphone and the reference sensor for source n, and $v_m(k)$ is the additive noise signal at the sensor m, assumed to be uncorrelated with both the source signals and the noise observed at the other sensors.

In far-field conditions, the relationship between TDOA and DOA can be solved easily with geometrical considerations. Therefore, for a generic pair of microphones with TDOA τ_n, DOA estimate is obtained as

$$\theta_n = \arcsin\left(\frac{\tau_n c}{d}\right) \tag{2}$$

where c is the speed of sound and d the distance between microphones.

The SRP relies on maximizing the power output of a beamformer. Broadband SRP operates in frequency-domain on a block-by-block basis. Consider a time-domain block of L samples. Beamforming can be seen as a filtered combination of the delayed signals, and the frequency-domain output of a generic beamformer in matrix notation for frequency f can be written as

$$Y(f) = \mathbf{W}^H(f)\mathbf{X}(f) \tag{3}$$

where $\mathbf{X} = [X_1(f)X_2(f)\dots X_M(f)]^T$, $Y(f)$ and $X_m(f)$ are the FFT of the signals, f is the frequency bin index, $\mathbf{W}(f) = [W_1(f)\ W_2(f)\dots W_M(f)]^T$ is the frequency vector of the beamformer weights for steering and filtering the data, and the superscript H represents the Hermitian (complex conjugate) transpose. The power spectral density of the beamformer output is given by

$$\begin{aligned} P(f) &= E[|Y(f)|^2] = \mathbf{W}^H(f)E[\mathbf{X}(f)\mathbf{X}^H(f)]\mathbf{W}(f) \\ &= \mathbf{W}^H(f)\mathbf{\Phi}(f)\mathbf{W}(f) \end{aligned} \tag{4}$$

where $\mathbf{\Phi}(f)$ is the cross-spectral density matrix and $E[\cdot]$ denotes mathematical expectation.

A. Narrowband SRP

In this section, we describe the algorithms of SRP, SRP-PHAT, MVDR and MUSIC.

The conventional SRP [1] consists in delaying and summing the block signals, and it can be written as

$$P_{\text{SRP}}(f,\tau) = \mathbf{A}^H(f,\tau)\mathbf{\Phi}(f)\mathbf{A}(f,\tau) \tag{5}$$

where $\mathbf{W}_{\text{SRP}}(f) = \mathbf{A}(f,\tau)$ is the steering vector corresponding to a given direction. We have introduced the dependence on the TDOA τ variable, and the equation (2) can be used for the TDOA-DOA transformation.

The SRP-PHAT [3] consists in applying the weighting function that divides the spectrum by its magnitude

$$P_{\text{SRP-PHAT}}(f,\tau) = \mathbf{A}^H(f,\tau)(\mathbf{\Phi}(f) \div |\mathbf{\Phi}(f)|)\mathbf{A}(f,\tau) \tag{6}$$

where $\div$ denotes element-wise division. Thus, PHAT filter simply discards the magnitude and only keeps the phase of $\mathbf{\Phi}$ for computing the steered responses.

The SRP with MVDR filter [4] relies on the solution of the minimization problem

$$\underset{\mathbf{W}(f)}{\text{argmin}} \mathbf{W}^H(f)\mathbf{\Phi}(f)\mathbf{W}(f) \quad \text{s.t.} \quad \mathbf{W}^H(f)\mathbf{A}(f,\tau) = 1. \tag{7}$$

The aim is to minimize the energy of noise and sources coming from different directions, while keeping a fixed gain on the desired direction. Solving (7) using the method of Lagrange multipliers, we can write

$$\mathbf{W}_{\text{MVDR}}(f) = \frac{\mathbf{\Phi}^{-1}(f)\mathbf{A}(f,\tau)}{\mathbf{A}^H(f,\tau)\mathbf{\Phi}^{-1}(f)\mathbf{A}(f,\tau)}. \tag{8}$$

In real applications, the inverse of the cross-spectral density matrix can be calculated using the Moore-Penrose pseudoinverse [11], defined as $\mathbf{\Phi}^+ = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^H$, where $\mathbf{\Phi} = \mathbf{U}\mathbf{S}\mathbf{V}^H$ is the singular value decomposition of the matrix $\mathbf{\Phi}$. Moreover, if $\mathbf{\Phi}$ is ill-conditioned, the spatial spectrum could be deteriorated by steering vector errors and finite sample effect [12]. Therefore, a diagonal loading (DL) [13] method is adopted to calculate the inverse matrix in a stable way. The power spectrum of the beamformer output with MVDR filter and DL becomes

$$P_{\text{MVDR}}(f,\tau) = \frac{1}{\mathbf{A}^H(f,\tau)(\mathbf{\Phi}(f) + \mu\mathbf{I})^+\mathbf{A}(f,\tau)} \tag{9}$$

where $\mathbf{I}$ is the identity matrix and $\mu = \frac{1}{L}\text{trace}[\mathbf{\Phi}(f)]\Delta$ is the loading level, where Δ is the normalized loading constant.

The MUSIC algorithm [6] is based on an eigen subspace decomposition method, and it exploits the orthogonality between signal and noise subspaces. By performing the eigenvalue decomposition of the cross-spectral density matrix, we obtain $\mathbf{\Phi} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M]$ is the square $M \times M$ matrix whose $\mathbf{u}_m$ is the mth eigenvector and $\mathbf{\Lambda}$ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues. MUSIC assumes that the N eigenvectors, which correspond to the N largest eigenvalues, span the signal subspace, and the remaining $M - N$ eigenvectors, which correspond to the zero eigenvalue, span the noise subspace. The subspace orthogonality property leads us to define the power pseudo-spectrum

$$P_{\text{MUSIC}}(f,\tau) = \frac{1}{\mathbf{A}^H(f,\tau)\mathbf{G}(f)\mathbf{G}^H(f)\mathbf{A}(f,\tau)} \tag{10}$$

where $\mathbf{G}(f)$ is the $M \times (M - N)$ matrix containing the eigenvectors corresponding to the noise subspace. MUSIC requires the analysis of eigenvalues for estimation of source number and it can be applied for localization in case of $N \leq M$.

B. Normalized Arithmetic Mean

The proposed incoherent averaging model is based on a normalized arithmetic mean (NAM), and it aims to mitigate the effect of incorrect response power estimation due to the variations of the SNR at each frequency and the GM problem. The goal is to obtain a SRP spectrum in which each frequency gives the same contribution to the final result, and this is achieved by implementing a normalization on power spectrum,

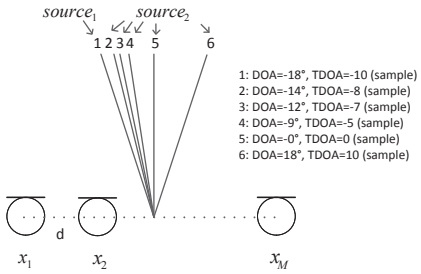

Fig. 1. The ULA and DOAs source position used in the simulated experiments.

by imposing a constraint for the values to be in the range $[0, 1]$. Thus, the NAM can be written as

$$P_{\text{NAM}}(\tau) = \sum_{f=0}^{L-1} \frac{P(f,\tau)}{\max\limits_{\tau'}[\mathbf{P}_{\tau'}(f)]} \tag{11}$$

where $\mathbf{P}_{\tau'}(f) = [P(f, -\tau_{max}), \ldots, P(f, \tau_{max})]$ is the vector of the power for all the desired direction ($\tau_{max} = df_s/c$ is maximum TDOA in samples for distance d and sampling frequency f_s) and $\max[\cdot]$ denotes the maximum value.

NAM is effective when used in combination with SRP, MVDR and MUSIC, but not with PHAT, which already provides a spectrum normalization and thus optimally performs with AM. We want to remark the difference between PHAT, which is a prefilter that sets all magnitude values to 1 on $\mathbf{\Phi}$ and only keeps the phase, and the novel approach, which is a postfilter on the narrowband power spectrum. Therefore, the proposed NAM allows to work on a full matrix $\mathbf{\Phi}$ for optimal performance of high resolution MVDR and MUSIC methods. Note that using MVDR and MUSIC with the PHAT pre-weighting means keeping only the phase of the cross-spectral density matrix for computing the steered response, thus reducing the benefits of the high resolution in low SNR conditions.

Finally, the values corresponding to the principal N peaks of the broadband steered power $\mathbf{P}^{\tau'}_{\text{NAM}} = [P_{\text{NAM}}(-\tau_{max}), \ldots, P_{\text{NAM}}(\tau_{max})]$ (in practice, those peaks which are above a given threshold) allow the TDOAs estimation of the N acoustic sources

$$\widehat{\tau}_n = \arg(\text{local})\max_{\tau'}[\mathbf{P}^{\tau'}_{\text{NAM}}] \quad n = 1, 2, \ldots, N. \tag{12}$$

The DOAs of sources on the array can be calculated using the equation (2) with the values estimate in (12).

III. Experiments

In this section, experiments on simulated data and a validation in a real-world scenario are reported.

For simulated data experiments, three uniform linear array (ULA) sizes have been used: a small array (3 microphones), a medium array (8 microphones), and a large array (24 microphones). For each array size, five tests have been performed to evaluate and compare the broadband SRP algorithms. A set of 50 Monte Carlo simulations with two motor vehicle signals have been used in different DOA positions. Figure 1 shows the considered setup. The first source is always positioned in

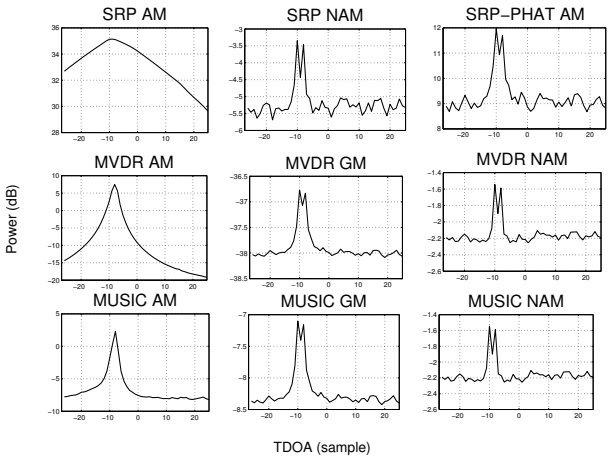

Fig. 2. Comparison of the power spectrum in a specific block with two sources in position 1 and 2, and an ULA of 8 microphones. Note that the broadband SRP with the proposed NAM provides an high spatial resolution, and an effective estimation of sources (the two power peaks are clearly visible).

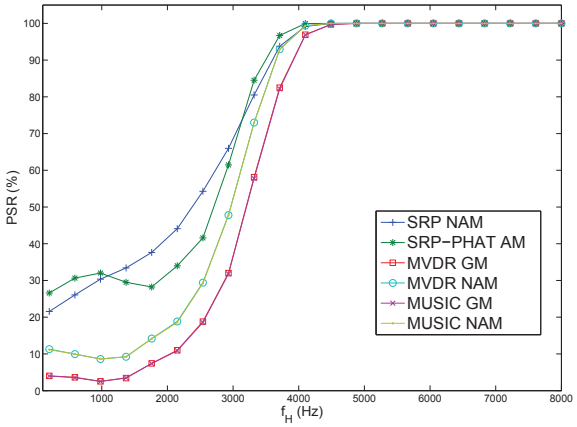

Fig. 3. Comparison of performance with variable bandwidth of two WGN signals in position 1 and 2, and an ULA of 8 microphones. The SNR was 20 dB and f_L was set to 100 Hz.

1, while the second source is positioned at increasing angular distances (positions 2, 3, 4, 5 and 6). The sampling frequency was 44.1 kHz, the signal block size was set to 2048 samples, and an Hann analysis window was used. The distance between microphones was 0.25 m. The normalized loading constant for MVDR was set to 0.001. The tests were conducted with different SNR levels, obtained by adding mutually independent white Gaussian noise (WGN) to each channel. We compare the performances of SRP with AM and NAM, SRP-PHAT AM, and MVDR and MUSIC with AM, GM and NAM. Table I shows the comparison of the performances, reporting the percentage success rate (PSR) obtained by dividing the number of correct DOA estimations for both sources in each block by the total number of analysis block. The power spectrum of a specific block is reported in Figure 2. The experimental results demonstrate that broadband SRP, with NAM averaging models, can be used as a high resolution method. NAM improves performance for SRP, MVDR and MUSIC. Moreover, we observe that MVDR and MUSIC has the same performance

TABLE I
COMPARISON OF PSR (%) FOR SRP WITH AM AND NAM, SRP-PHAT AM, AND MVDR AND MUSIC WITH AM, GM AND NAM.

3 MICROPHONES									
SNR (dB)	SRP AM	SPR NAM	SRP-PHAT AM	MVDR AM	MVDR GM	MVDR NAM	MUSIC AM	MUSIC GM	MUSIC NAM
-10	0.00	**3.40**	3.58	4.42	4.53	**4.16**	4.64	4.55	**4.16**
-5	0.00	**6.52**	6.58	5.75	8.41	**8.01**	7.84	8.39	**8.03**
0	0.00	**10.41**	10.73	9.90	12.63	**12.85**	12.34	12.67	**12.91**
5	4.40	**15.56**	15.24	13.19	17.37	**17.98**	16.17	17.43	**18.00**
10	3.37	**22.31**	21.07	15.31	21.67	**23.71**	20.57	21.71	**23.68**
15	3.06	**38.94**	34.91	18.94	35.47	**41.12**	28.61	35.50	**41.10**
20	3.71	**72.50**	67.29	21.59	60.18	**69.02**	35.69	60.13	**69.22**
8 MICROPHONES									
SNR (dB)	SRP AM	SPR NAM	SRP-PHAT AM	MVDR AM	MVDR GM	MVDR NAM	MUSIC AM	MUSIC GM	MUSIC NAM
-10	0.00	**9.50**	11.18	6.98	14.11	**13.38**	13.09	14.10	**13.37**
-5	2.59	**21.79**	22.19	3.90	22.09	**23.70**	8.52	22.07	**23.69**
0	10.06	**29.94**	30.19	18.19	34.51	**36.96**	29.42	34.47	**36.96**
5	10.88	**58.70**	53.94	24.06	58.84	**67.02**	44.76	58.86	**67.03**
10	11.79	**91.63**	82.88	29.21	78.08	**88.25**	58.65	78.09	**88.25**
15	11.57	**99.79**	97.06	30.57	90.09	**99.13**	62.69	90.11	**99.13**
20	11.33	**100.00**	100.00	30.19	99.69	**100.00**	62.69	99.70	**100.00**
24 MICROPHONES									
SNR (dB)	SRP AM	SPR NAM	SRP-PHAT AM	MVDR AM	MVDR GM	MVDR NAM	MUSIC AM	MUSIC GM	MUSIC NAM
-10	16.97	**34.81**	35.17	4.98	36.43	**42.16**	10.17	36.41	**42.16**
-5	36.04	**73.21**	78.74	4.93	77.97	**89.15**	12.48	77.97	**89.15**
0	62.93	**94.46**	96.37	44.85	98.74	**99.64**	80.06	98.74	**99.64**
5	63.63	**99.85**	99.94	48.52	99.94	**100.00**	84.49	99.94	**100.00**
10	63.97	**99.99**	100.00	48.95	100.00	**100.00**	84.86	100.00	**100.00**
15	64.24	**100.00**	100.00	48.97	100.00	**100.00**	82.59	100.00	**100.00**
20	64.29	**100.00**	100.00	49.98	100.00	**100.00**	82.15	100.00	**100.00**

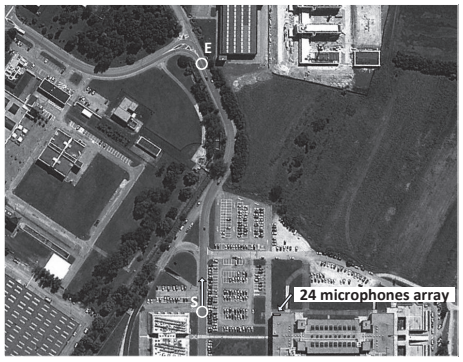

Fig. 4. The map with the position of the array and of point S and E.

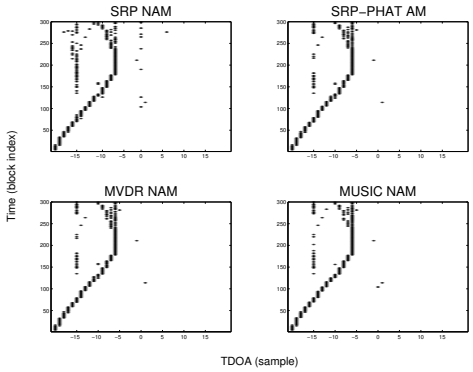

Fig. 5. The estimated $\widehat{\tau}_n$ for the motorcycle moving from point S to E.

with better accuracy in low SNR conditions if compared with SRP NAM and SRP-PHAT AM. Note that the performance of SRP NAM is similar to MVDR NAM and MUSIC NAM up to 5 dB SNR.

A second simulated experiment is reported to evaluate the NAM performance with variable bandwidth signals. Two WGN signals, positioned in 1 and 2 with an ULA of 8 microphones, are processed with a bandpass filter $[f_L, f_H]$, where f_L and f_H are the lower and upper frequency limit respectively. Comparison of broadband SRP is depicted in Figure 3 for a SNR of 20 dB using 50-run trials for each bandwidth. All power responses are characterized by a decrease in performance when the sources become narrowband. Therefore, a fusion using an optimal range of frequencies is desirable in these cases. As can be observed, NAM performs better than GM, and SRP NAM is the most effective, except for bandwidths below 1000 Hz and in the 3000-4500 Hz range, in which the SRP-PHAT AM has a greater PSR.

In order to evaluate the proposed NAM, a validation in a real-world scenario is reported. An ULA of 24 microphones has been installed on the roof on the University building. The microphone distance was 0.15 m, and a sample rate of 48 kHz has been used. The DOA estimation of a moving motorcycle is considered. Figure 4 shows the map with the position of the array and the street that the motorcycle has traveled, from point S to E. The array position is orthogonal to the street at point S. The distance of the array from point S is 68 m, and 250 m from point E. In Figure 5, we can see the TDOAs estimated for SRP, MVDR and MUSIC with the proposed NAM method and SRP-PHAT with AM. All methods provide the correct localization of the source from point S to E in the open space at a large distance. We can also note in the figure, a second source on the left of the motorcycle trajectory, from time block index 150 and with a TDOA of -15 samples.

IV. CONCLUSIONS

Fusing narrowband power of each frequency bin is a crucial step for accuracy broadband steered response power. An incoherent combination based on NAM is proposed to mitigate the effect of incorrect narrowband power spectrum due to SNR variability at each frequency and to mitigate the problem due to GM fusion. NAM consists on applying a postfilter on each narrowband steered response power before computing fusion. Experimental results demonstrate the improvement provided by this solution for SRP, MVDR and MUSIC. Comparison of broadband power responses shows that SRP with NAM is suitable for high resolution.

REFERENCES

[1] M. S. Bartlett, "Smoothing periodograms from time-series with continuous spectra," *Nature*, vol. 161, pp. 686–687, 1948.

[2] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[3] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, *Microphone Arrays: Signal Processing Techniques and Applications*. Springer, 2001, ch. Robust localization in reverberant rooms.

[4] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[5] M. R. Azimi-Sadjadi, A. Pezeshki, and N. Roseveare, "Wideband DOA estimation algorithms for multiple moving sources using unattended acoustic sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 4, pp. 1585–1599, 2008.

[6] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[7] M. Wax and T. Kailath, "Spatio-temporal spectral analysis by eigenstructure methods," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 4, pp. 817–827, 1984.

[8] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 4, pp. 823–831, 1985.

[9] E. Di Claudio and R. Parisi, "WAVES: Weighted average of signal subspaces for robust wideband direction finding," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2179–2190, 2001.

[10] Y. Yoon, L. M. Kaplan, and J. H. McClellan, "TOPS: New DOA estimator for wideband signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 791–802, 2006.

[11] C. Pan, J. Chen, and J. Benesty, "Performance study of the MVDR beamformer as a function of the source incidence angle," *IEEE/ACM Transactions on Audio, Speech, Language Processing*, vol. 22, pp. 67–79, 2014.

[12] Y. L. Chen and J.-H. Lee, "Finite data performance analysis of MVDR antenna array beamformers with diagonal loading," *Progress In Electromagnetics Research*, vol. 134, pp. 475–507, 2013.

[13] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.